

Frederick Albert Matsen IV

Computational Biology Program
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N., S2-105
Mail stop: S2-140
Seattle, WA 98109-1024

matsen@fredhutch.org
<http://matsen.fredhutch.org/>

Positions

- Investigator**, Howard Hughes Medical Institute.
October 2021 – Present
- Professor**, Fred Hutchinson Cancer Research Center, Seattle.
April 2020 – Present
- Associate Member**, Fred Hutchinson Cancer Research Center, Seattle.
April 2015 – March 2020
- Affiliate Associate Professor**, Genome Sciences, University of Washington, Seattle.
March 2017 – Present
- Affiliate Associate Professor**, Statistics, University of Washington, Seattle.
June 2016 – Present
- Assistant Member**, Fred Hutchinson Cancer Research Center, Seattle.
August 2010 – March 2015
- Miller Postdoctoral Scholar**, UC Berkeley, Berkeley, California.
July 2007 – July 2010
- Allan Wilson Postdoctoral Scholar**, University of Canterbury, Christchurch, New Zealand.
July 2006 – June 2007

Education

- Harvard University**, Cambridge, MA.
Ph.D. in Mathematics, June 2006.
- Université de Paris 13**, Paris, France.
D.E.A. in Mathematics, June 2001.
- Stanford University**, Stanford, CA.
B.S. in Mathematics with honors, June 2000.

Grants:

- HHMI Investigator Award** (2021/10/01 – 2028/09/31)
- R01 AI162611** (2021/07/01 – 2026/06/31)
Fast and flexible Bayesian phylogenetics via modern machine learning.
- R01 AI146028** (2019/06/01 – 2024/05/31)
Blending deep learning with probabilistic mechanistic models to predict and understand the evolution and function of adaptive immune receptors.
- HHMI-Simons Faculty Scholar Award** (11/01/2016 – 11/01/2021)
- NSF Award 1564137** (2016/07/01 – 2020/06/30)
Enabling phylogenetic inference for modern data sets.
Co-PI with Vladimir Minin.
- R01 GM1132461** (2014/08/01 – 2019/08/01)
Leveraging deep sequencing data to understand antibody maturation.
- NSF Award 1223057** (2012/09/01 – 2017/08/31)
New theorems and algorithms for comprehensive analysis of metagenomic data via statistical phylogenetics.
Co-PI with Aaron Darling.

Frederick Albert Matsen IV

R01 HG005966 (2010/09/27 – 2013/06/30)

Novel Computational Tools for Studying the Human Microbiome.

Co-PI with David Fredricks.

ITHS pilot grant (2014/02/01 – 2015/02/01)

Inferring statistical models of antibody somatic hypermutation from high-throughput sequencing data

CFAR New Investigator Award (2014/01/01 – 2016/01/01)

Machine learning, informed by structural knowledge, to expand our knowledge of APOBEC3 specificity and improve hypermutation inference

Grant support:
co-investigator

R01 AI138709 (2018/02/07 – 2023/02/07)

Characterizing the Broad Antibody Response to HIV Superinfection

PI: Julie Overbaugh.

Our role: our group will design and perform all the B cell receptor sequence analysis for this project.

R01 AI120961 (2016/01/01 – 2020/12/31)

Defining the Infant Immune Response to HIV.

PI: Julie Overbaugh.

Our role: our group will design and perform all the B cell receptor sequence analysis for this project.

Bill and Melinda Gates Foundation OPP1110049 (2014/08/31 – 2016/07/31)

Biostatistical, Computational Biology, and Mathematical Modeling for the Assessment of Immune Correlates of Protection in the HVTN 701 and 702 Efficacy Trials in South Africa.

PI: Peter Gilbert.

Our role: Improve methods for reconstruction of HIV molecular sequences which found infection using later time point samples.

R37 AI038518-16A2 (2010/03/01 – 2015/02/28)

Early and Reinfection in High Risk Women.

PI: Julie Overbaugh.

Our role: our group has designed and performed all the bioinformatics for this project.

R21 AI104449-01 (2013/02/15 – 2015/01/31)

PrEP Exposure during HIV-1 Infection and the Risk of Low-Frequency ARV Resistance.

PI: Dara Lehman.

Our role: our group has designed and performed all the bioinformatics for this project.

U54GM111274-01 (2014/09/12 – 2019/06/30)

Center for Statistics and Quantitative Infectious Diseases.

PI: Elizabeth Halloran.

Our role: design and implement computational phylogenetic algorithms for pathogen sequence data.

Awards

HHMI-Simons faculty scholar (November 2016 – October 2021)

Research support for five years.

Miller fellow (July 2007 – July 2010)

Support for three years of postdoctoral research at UC Berkeley.

Allan Wilson postdoctoral fellow (July 2006 – June 2007)

Support for one year of postdoctoral research.

NSF graduate research fellow (July 2003 – June 2006)

Support for three years of study.

Harvard University certificate of distinction in teaching (Fall 2004)

For teaching second semester calculus.

Frederick Albert Matsen IV

Jean D. Valpine fellowship (September 2001 – June 2003)

Support for two years of study.

Fulbright fellowship (July 2000 – June 2001)

To study mathematics in Paris.

Research interests

phylogenetics, innate and adaptive immunity, infectious disease, human microbiome, Markov processes, coalescent theory, molecular evolution, macroevolutionary models, discrete geometry, combinatorics, Bayesian statistics, bioinformatics.

Alumni

Masters and PhD students

Amrit Dhar	2016-2019	Data Scientist, Google
Jean Feng	2016-2020	Assistant professor (tenure track), UCSF
Hassan Nasif	2020-2022	Data Scientist, Microsoft

Postdocs

Vu Dinh	2015-2017	Assistant professor (tenure track), University of Delaware
David Shaw	2015-2018	Data Scientist, Aurora
Julia Fukuyama	2017-2018	Assistant professor (tenure track), University of Indiana
Chris Whidden	2013-2018	Assistant professor (tenure track), Dalhousie University
Cheng Zhang	2016-2019	Assistant professor (tenure track), Peking University
Michael Karcher	2019-2021	Assistant professor (tenure track), Muhlenberg College

Activities

PhD committee member: Miguel Parades, since 2023.

PhD committee member: Marlin Figgins, Alex Willcox and Tim Yu, since 2022.

PhD committee member: Malika Hale, William Hannon, Zachary Montague, and Elena Violeta Romero, since 2021.

PhD committee member: Frances Welsh and Zak Yaffe, since 2020.

PhD committee member: Caelan Radford, since 2019.

Conference Co-organizer: *Advances in Immune Repertoire Measurement and Analysis*, Chan-Zuckerberg Biohub, January 2019.

PhD committee member: Allison Black, John Huddleston and Anna Minkina, since 2018.

PhD committee member: Sarah Hilton and Sydney Bell, since 2016.

Invited session organizer: *Bernoulli Society World Congress in Probability and Statistics*, July 2016.

Mentor: Summer Undergraduate Research Program, Summer 2016.

Mentor: Summer High School Internship Program, Summer 2014-6.

Organizer: Spotlight session on next generation phylogenetic inference, *Joint Evolution Meetings*, June 2016.

Leader: Tools and Resources Working Group, Adaptive Immune Receptor Repertoire Community. (<http://airr-community.org>)

Founder and organizer: *B-T.CR* online immunogenetics discussion forum (<http://b-t.cr/>), since February 2016.

Associate editor: *PLOS Computational Biology*, since December 2015.

Editor: Special issue of *Philosophical Transactions of the Royal Society B* on antibody dynamics, 2015.

Invited program committee member: *International Conference on Intelligent Systems for Molecular Biology (ISMB) / European Conference on Computational Biology (ECCB)*, 2009 and 2015.

Frederick Albert Matsen IV

Member: Statistical Genetics faculty group at the University of Washington, since October 2014.

Founder: *fredhutch.io* initiative to improve access to and education on scientific computing at the Fred Hutch (<http://fredhutch.io/>), since July 2014.

Founder and organizer: *Phylobabble* online phylogenetics discussion forum (<http://phylobabble.org/>), March 2014 – 2022.

Editorial board: *Systematic Biology*, since February 2013.

Host for professor on sabbatical: Vladimir Minin, academic year 2013-4.

Founder and organizer: *Phyloseminar* online phylogenetics seminar (<http://phyloseminar.org/>), since November 2009.

Contributor: *Biopython project* (<http://github.com/biopython/>), since December 2010.

Conference co-organizer: *Molecular Evolution and Phylogenetics*, Statistical and Applied Mathematical Sciences Institute (SAMSI), Spring 2009.

Paper and grant reviewer (selected list): *Annals of Human Genetics*, *Bioinformatics*, *Journal of Bioinformatics and Computational Biology*, *Bulletin of Mathematical Biology*, *ISMB/EECB 2015*, *Journal of Mathematical Biology*, *Journal of Molecular Evolution*, *Molecular Biology and Evolution*, *National Science Foundation*, *PLOS Pathogens*, *PNAS*, *STOC08*, *Systematic Biology*, *IEEE Transactions on Computational Biology and Bioinformatics*.

Invited and
conference
presentations

“Analyzing the space of TCRs using optimal transport” *Adaptive Immune Receptor Repertoire Community (AIRR-C) Annual Meeting*, December 2020.

“Learning, using, and extending variational distributions of phylogenetic trees” *Machine Learning for Computational Biology*, December 2019.

“Variational Bayesian Phylogenetic Inference” *Workshop in Biostatistics, Stanford University*, October 2019.

“Beyond random-walk Markov chain Monte Carlo for Bayesian phylogenetics” *Joint Statistical Meetings*, July 2018.

“Taking uncertainty and biology seriously when analyzing B cell receptor sequence data” *Boston University Computational Immunology Symposium*, June 2018.

“Beyond random-walk Markov chain Monte Carlo for Bayesian phylogenetics” *Harvard Biomathematics seminar series*, June 2018.

“Learning about antibody affinity maturation from sequence data using probabilistic models” *Penn Symposium on Mathematical & Computational Biology*, May 2017.

“Learning from immune repertoire sequences with probabilistic methods” *Stanford Computational and Systems Immunology Series*, August 2016.

“Consistency and convergence rate of phylogenetic inference via regularization” *Bernoulli Society World Congress in Probability and Statistics*, July 2016.

“The adaptive immune system: a grand and beautiful stochastic process” *University of Washington Statistics Seminar*, April 2016.

“Learning how antibodies are drafted and revised” *Simon Fraser University*, March 2016, *University of Southern California*, November 2015, *University of Pennsylvania*, November 2015, *National Bioterrorism and Countermeasures Center*, Frederick, MD, June 2015.

“Degree, curvature, and mixing of random walks on the phylogenetic subtree-prune-regraft graph” *Annual meeting of the Cladistics Society*, June 2015.

“Statistical inference of annotations and rearrangement groups and evolutionary constraints on B cell receptors” *Keystone Symposia on The Golden Anniversary of B Cell Discovery*, March 2015.

Frederick Albert Matsen IV

“Learning how antibodies are drafted and revised” *Center for Computational Biology Seminar Series*, UC Berkeley, April 2015.

“Leveraging the discrete and continuous structure of phylogenetic trees for the analysis of metagenomic data” *Joint Statistical Meetings*, August 2014.

“From the Ramayana to the Reverend Bayes: host defenses and zoonotic transmission of simian foamy virus” *Collège de France*, Paris, France, June 2013.

“Phylogenetics and the human microbiome” *LIRMM Molecular and Computational Biology Workshop*, Montpellier, France, June 2013.

“Analysis of metagenomes using phylogenetics” *American Society of Microbiology General Meeting*, Denver, CO, May 2013.

“Phylogenetics for environmental genomics” *Phycological Society of America Annual Meeting*, Seattle, WA, July 2011.

“Analysis and comparison of microbial sequence data via phylogenetic placements” *International Human Microbiome Congress*, Vancouver, BC, March 2011.

“Analysis and comparison of microbial sequence data via phylogenetic placements” *High-Throughput Biodiversity Research Using Eukaryotic Metagenetics*, NESCENT (Durham, NC), January 2011.

“The phylogenetic Kantorovich-Rubinstein metric for environmental DNA sequence samples” *Joint Statistical Meetings*, Vancouver, August 2010.

“From phytoplankta to rogue taxa: phylogenetics theory and practice” *Workshop in Biostatistics*, Stanford, February 2010.

“The phylogenetic geometry of rogue taxa” *New Mathematical Challenges from Molecular Biology and Genetics*, Banff International Research Station, Fall 2009.

“The phylogenetic geometry of rogue taxa” *Cyberinfrastructure for Phylogenetic Research (CIPRES) All Hands Meeting*, Berkeley, Fall 2009.

“Glimpses from the strange world of phylogenetic mixtures” *Future Directions in Phylogenetics*, Isaac Newton Institute for Mathematical Sciences, Fall 2008.

“Discrete harmonic analysis and the geometry of model-based phylogenetics” *Workshop on Interactions between Probability Theory and Computer Science*, Cornell, 2008.

Publications

Zak A Yaffe, Shilei Ding, Kevin Sung, Vrasha Chohan, Lorie Marchitto, Laura Doepker, Duncan Ralph, Ruth Nduati, Frederick A Matsen, Andrés Finzi, and Julie Overbaugh. Reconstruction of a polyclonal ADCC antibody repertoire from an HIV-1 Non-Transmitting mother. *iScience*, page 106762, April 2023.

Andrew F Magee, Michael D Karcher, Frederick A Matsen, IV, and Vladimir N Minin. How trustworthy is your tree? Bayesian phylogenetic effective sample size through the lens of Monte Carlo error. *arXiv*, September 2021.

Michael Karcher, Cheng Zhang, and Frederick A Matsen, IV. Variational Bayesian supertrees. *arXiv*, April 2021.

Christiaan Swanepoel, Mathieu Fourment, Xiang Ji, Hassan Nasif, Marc A Suchard, Frederick A Matsen, IV, and Alexei Drummond. TreeFlow: probabilistic programming and automatic differentiation for phylogenetics. *arXiv*, November 2022.

Caitlin I Stoddard, Kevin Sung, Zak A Yaffe, Haidyn Weight, Guillaume Beaudoin-Bussièrès, Jared Galloway, Soren Gantt, Judith Adhiambo, Emily R Begnel, Ednah Ojee, Jennifer Slyker, Dalton Wamalwa, John Kinuthia, Andrés Finzi, Frederick A Matsen, 4th, Dara A Lehman, and Julie Overbaugh. Elevated binding and functional antibody responses to SARS-CoV-2 in infants versus mothers. *Nat. Commun.*, 14(1):4864, August 2023.

Frederick Albert Matsen IV

Mathieu Fourment, Christiaan J Swanepoel, Jared G Galloway, Xiang Ji, Karthik Gan-gavarapu, Marc A Suchard, and Frederick A Matsen, Iv. Automatic differentiation is no panacea for phylogenetic gradient computation. *Genome Biol. Evol.*, 15(6), June 2023.

Magdalena L Russell, Noah Simon, Philip Bradley, and Frederick A Matsen, IV. Statistical inference reveals the role of length, GC content, and local sequence in V(D)J nucleotide trimming. *Elife*, 12:e85145, May 2023.

Timothy C Yu, Zorian T Thornton, William W Hannon, William S DeWitt, Caelan E Radford, Frederick A Matsen, 4th, and Jesse D Bloom. A biophysical model of viral escape from polyclonal antibodies. *Virus Evol*, 8(2):veac110, December 2022.

Branden J Olson, Stefan A Schattgen, Paul G Thomas, Philip Bradley, and Frederick A Matsen, IV. Comparing T cell receptor repertoires using optimal transport. *PLoS Comput. Biol.*, 18(12):e1010681, December 2022.

Frederick A Matsen IV and Peter L Ralph. Enabling inference for Context-Dependent models of mutation by bounding the propagation of dependency. *J. Comput. Biol.*, 29(8):802–824, August 2022.

Caitlin I Stoddard, Kevin Sung, Ednah Ojee, Judith Adhiambo, Emily R Begnel, Jennifer Slyker, Soren Gantt, Frederick A Matsen, John Kinuthia, Dalton Wamalwa, Julie Overbaugh, and Dara A Lehman. Distinct antibody responses to endemic coronaviruses pre- and Post-SARS-CoV-2 infection in kenyan infants and mothers. *Viruses*, 14(7):1517, July 2022.

Alexandra C Willcox, Kevin Sung, Meghan E Garrett, Jared G Galloway, Jesse H Erasmus, Jennifer K Logue, David W Hawman, Helen Y Chu, Kim J Hasenkrug, Deborah H Fuller, Frederick A Matsen, Iv, and Julie Overbaugh. Detailed analysis of antibody responses to SARS-CoV-2 vaccination and infection in macaques. *PLoS Pathog.*, 18(4):e1010155, April 2022.

Duncan K Ralph and Frederick A Matsen, 4th. Inference of B cell clonal families using heavy/light chain pairing information. *PLoS Comput. Biol.*, 18(11):e1010723, November 2022.

Magdalena L Russell, Aisha Souquette, David M Levine, Stefan A Schattgen, E Kaitlynn Allen, Guillermina Kuan, Noah Simon, Angel Balmaseda, Aubree Gordon, Paul G Thomas, Frederick A Matsen, IV, and Philip Bradley. Combining genotypes and T cell receptor distributions to infer genetic loci determining V(D)J recombination probabilities. *Elife*, 11:e73475, March 2022.

Meghan E Garrett, Jared G Galloway, Caitlin Wolf, Jennifer K Logue, Nicholas Franko, Helen Y Chu, Frederick A Matsen, IV, and Julie M Overbaugh. Comprehensive characterization of the antibody responses to SARS-CoV-2 spike protein finds additional vaccine-induced epitopes beyond those for mild infection. *Elife*, 11:e73490, January 2022.

Hannah L Itell, Haidyn Weight, Carolyn S Fish, Jennifer K Logue, Nicholas Franko, Caitlin R Wolf, Denise J McCulloch, Jared Galloway, Frederick A Matsen, 4th, Helen Y Chu, and Julie Overbaugh. SARS-CoV-2 antibody binding and neutralization in dried blood spot eluates and paired plasma. *Microbiol Spectr*, page e0129821, October 2021.

Mackenzie M Shipley, Vidya Mangala Prasad, Laura E Doepker, Adam S Dingens, Duncan K Ralph, Elias Harkins, Amrit Dhar, Dana Arenz, Vrasha Chohan, Haidyn Weight, Kishor Mandaliya, Jesse D Bloom, Frederick Matsen, IV, Kelly K Lee, and Julie M Overbaugh. Functional development of a v3/glycan-specific broadly neutralizing antibody isolated from a case of HIV superinfection. *Elife*, 10:e68110, July 2021.

Laura E Doepker, Sonja Danon, Elias Harkins, Duncan K Ralph, Zak Yaffe, Meghan E Garrett, Amrit Dhar, Cassia Wagner, Megan M Stumpf, Dana Arenz, James A Williams, Walter Jaoko, Kishor Mandaliya, Kelly K Lee, Frederick A Matsen, 4th, and Julie M Overbaugh. Development of antibody-dependent cell cytotoxicity function in HIV-1 antibodies. *Elife*, 10, January 2021.

Meghan E Garrett, Jared Galloway, Helen Y Chu, Hannah L Itell, Caitlin I Stoddard, Caitlin R Wolf, Jennifer K Logue, Dylan McDonald, Haidyn Weight, Frederick A Matsen, and Julie Overbaugh. High-resolution profiling of pathways of escape for SARS-CoV-2 spike-binding antibodies. *Cell*, May 2021.

Caitlin I Stoddard, Jared Galloway, Helen Y Chu, Mackenzie M Shipley, Kevin Sung, Hannah L Itell, Caitlin R Wolf, Jennifer K Logue, Ariana Magedson, Meghan E Garrett, Katharine H D Crawford, Uri Laserson, Frederick A Matsen, and Julie Overbaugh. Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs. *Cell Rep.*, 35(8):109164, May 2021.

Jean Feng, William S DeWitt, III, Aaron McKenna, Noah Simon, Amy D Willis, and Frederick A Matsen, IV. Estimation of cell lineage trees by maximum-likelihood phylogenetics. *Annals of Applied Statistics*, 15(1):343–362, March 2021.

Mark D Pankau, Daniel B Reeves, Elias Harkins, Keshet Ronen, Walter Jaoko, Kishor Mandalia, Susan M Graham, R Scott McClelland, Frederick A Matsen, IV, Joshua T Schiffer, Julie Overbaugh, and Dara A Lehman. Dynamics of HIV DNA reservoir seeding in a cohort of superinfected kenyan women. *PLoS Pathog.*, 16(2):e1008286, February 2020.

Duncan K Ralph and Frederick A Matsen, IV. Using B cell receptor lineage structures to predict affinity. *PLoS Comput. Biol.*, 16(11):e1008391, November 2020.

Amrit Dhar, Duncan K Ralph, Vladimir N Minin, and Frederick A Matsen IV. A bayesian phylogenetic hidden markov model for B cell receptor sequence analysis. *PLoS Comput. Biol.*, 16(8):e1008030, August 2020.

Cheng Zhang, Vu Dinh, and Frederick A Matsen. Nonbifurcating phylogenetic tree inference via the adaptive LASSO. *J. Am. Stat. Assoc.*, pages 1–16, June 2020.

Julia Fukuyama, Branden J Olson, and Frederick A Matsen, 4th. Lack of evidence for a substantial rate of templated mutagenesis in B cell diversification. *J. Immunol.*, July 2020.

Chris Whidden, Brian C Claywell, Thayer Fisher, Andrew F Magee, Mathieu Fourment, and Frederick A Matsen, IV. Systematic exploration of the high likelihood set of phylogenetic tree topologies. *Syst. Biol.*, 69(2):280–293, March 2020.

Mathieu Fourment, Andrew F Magee, Chris Whidden, Arman Bilge, Frederick A Matsen, IV, and Vladimir N Minin. 19 dubious ways to compute the marginal likelihood of a phylogenetic tree topology. *Syst. Biol.*, 69(2):209–220, March 2020.

Laura E Doepker, Cassandra A Simonich, Duncan Ralph, Mackenzie M Shipley, Meghan Garrett, Theodore Gobillot, Vladimir Vigdorovich, D Noah Sather, Ruth Nduati, Frederick A Matsen, 4th, and Julie M Overbaugh. Diversity and function of maternal HIV-1-specific antibodies at the time of vertical transmission. *J. Virol.*, February 2020.

William Lees, Christian E Busse, Martin Corcoran, Mats Ohlin, Cathrine Scheepers, Frederick A Matsen, Gur Yaari, Corey T Watson, AIRR Community, Andrew Collins, and Adrian J Shepherd. OGRDB: a reference database of inferred immune receptor genes. *Nucleic Acids Res.*, 48(D1):D964–D970, January 2020.

Jean Feng, David A Shaw, Vladimir N Minin, Noah Simon, and Frederick A Matsen, IV. Survival analysis of DNA mutation motifs with penalized proportional hazards. *Ann. Appl. Stat.*, 13(2):1268–1294, June 2019.

Kristian Davidsen, Branden J Olson, William S DeWitt, III, Jean Feng, Elias Harkins, Philip Bradley, and Frederick A Matsen, IV. Deep generative models for T cell receptor protein sequences. *Elife*, 8:e46935, September 2019.

Cassandra A Simonich, Laura Doepker, Duncan Ralph, James A Williams, Amrit Dhar, Zak Yaffe, Lauren Gentles, Christopher T Small, Brian Oliver, Vladimir Vigdorovich, Vidya Mangala Prasad, Ruth Nduati, D Noah Sather, Kelly K Lee, Frederick A Matsen IV, and Julie Overbaugh. Kappa chain maturation helps drive rapid development of an infant HIV-1 broadly neutralizing antibody lineage. *Nat. Commun.*, 10(1):2190, May 2019.

David A Shaw, Vu C Dinh, and Frederick A Matsen IV. Joint maximum likelihood of phylogeny and ancestral states is not consistent. *Mol. Biol. Evol.*, May 2019.

Lam Si Tung Ho, Vu Dinh, Frederick A Matsen, 4th, and Marc A Suchard. On the convergence of the maximum likelihood estimator for the transition rate under a 2-state symmetric model. *J. Math. Biol.*, 80(4):1119–1138, March 2020.

Cheng Zhang and Frederick A Matsen, IV. Variational bayesian phylogenetic inference. In *International Conference on Learning Representations (ICLR)*, 2019.

Cheng Zhang and Frederick A Matsen IV. Generalizing tree probability estimation via Bayesian networks. In S Bengio, H Wallach, H Larochelle, K Grauman, N Cesa-Bianchi, and R Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 1449–1458. Curran Associates, Inc., 2018. <http://papers.nips.cc/paper/7418-generalizing-tree-pdf>.

Amrit Dhar, Kristian Davidsen, Frederick A Matsen, 4th, and Vladimir N Minin. Predicting B cell receptor substitution profiles using public repertoire data. *PLoS Comput. Biol.*, 14(10):e1006388, October 2018.

Kristian Davidsen and Frederick A Matsen. Benchmarking tree and ancestral sequence inference for B cell receptor sequences. *Front. Immunol.*, 9:2451, 2018.

William S DeWitt, Anajane Smith, Gary Schoch, John A Hansen, Frederick A Matsen, and Philip Bradley. Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *Elife*, 7:e38358, August 2018.

Vu Dinh, Lam Si Tung Ho, Marc A Suchard, and Frederick A Matsen IV. Consistency and convergence rate of phylogenetic inference via regularization. *Ann. Stat.*, 46(4):1481–1512, August 2018.

Branden J Olson and Frederick A Matsen IV. The Bayesian optimist’s guide to adaptive immune receptor repertoire analysis. *Immunol. Rev.*, 284(1):148–166, June 2018.

William S DeWitt, Luka Mesin, Gabriel D Victora, Vladimir N Minin, and Frederick A Matsen IV. Using genotype abundance to improve phylogenetic inference. *Mol. Biol. Evol.*, February 2018.

F A Matsen IV, S C Billey, A Kas, and M Konvalinka. Tanglegrams: A reduction tool for mathematical phylogenetics. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 15(1):343–349, January 2018.

C Whidden and F Matsen IV. Efficiently inferring pairwise subtree prune-and-regraft adjacencies between phylogenetic trees. In *2018 Proceedings of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, Proceedings, pages 77–91. Society for Industrial and Applied Mathematics, January 2018.

Brian C Claywell, Vu Dinh, Mathieu Fourment, Connor O McCoy, and Frederick A Matsen IV. A surrogate function for one-dimensional phylogenetic likelihoods. *Mol. Biol. Evol.*, 35(1):242–246, January 2018.

Vu Dinh and Frederick A Matsen IV. The shape of the one-dimensional phylogenetic likelihood function. *Ann. Appl. Probab.*, 27(3):1646–1677, June 2017.

Vu Dinh, Arman Bilge, Cheng Zhang, and Frederick A Matsen IV. Probabilistic path Hamiltonian Monte Carlo. In *International Conference on Machine Learning*, pages 1009–1018. PMLR, 17 July 2017.

Vu Dinh, Aaron E Darling, and Frederick A Matsen IV. Online Bayesian phylogenetic inference: theoretical foundations via Sequential Monte Carlo. *Syst. Biol.*, December 2017.

Mathieu Fourment, Brian C Claywell, Vu Dinh, Connor McCoy, Frederick A Matsen IV, and Aaron E Darling. Effective online Bayesian phylogenetics via Sequential Monte Carlo with guided proposals. *Syst. Biol.*, November 2017.

Chris Whidden and Frederick A Matsen IV. Ricci–Ollivier curvature of the rooted phylogenetic subtree–prune– regraft graph. *Theor. Comput. Sci.*, 699(Supplement C):1–20, November 2017.

Felix Breden, Eline T Luning Prak, Bjoern Peters, Florian Rubelt, Chaim A Schramm, Christian E Busse, Jason A Vander Heiden, Scott Christley, Syed Ahmad Chan Bukhari, Adrian Thorogood, Frederick A Matsen IV, Yariv Wine, Uri Laserson, David Klatzmann, Daniel C Douek, Marie-Paule Lefranc, Andrew M Collins, Tania Bubela, Steven H Kleinstein, Corey T Watson, Lindsay G Cowell, Jamie K Scott, and Thomas B Kepler. Reproducibility and reuse of adaptive immune receptor repertoire data. *Front. Immunol.*, 8:1418, 2017.

Corey T Watson, Frederick A Matsen IV, Katherine J L Jackson, Ali Bashir, Melissa Laird Smith, Jacob Glanville, Felix Breden, Steven H Kleinstein, Andrew M Collins, and Christian E Busse. Comment on “A database of human immune receptor alleles recovered from population sequencing data”. *The Journal of Immunology*, 198(9):3371–3373, May 2017.

Alex Gavryushkin, Chris Whidden, and Frederick A Matsen IV. The combinatorics of discrete time-trees: theory and open problems. *J. Math. Biol.*, 29 July 2017.

Julie F Weis, Jared M Baeten, Connor O McCoy, Chris Warth, Deborah Donnell, Katherine K Thomas, Craig W Hendrix, Mark A Marzinke, Nelly Mugo, Frederick A Matsen IV, Connie Celum, Dara A Lehman, and for Partners PrEP Study Team. Preexposure prophylaxis-selected drug resistance decays rapidly after drug cessation. *AIDS*, 30(1):31, January 2016.

Chris Whidden and Frederick A Matsen IV. Chain reduction preserves the unrooted subtree Prune-and-Regraft distance. *arXiv*, 8 November 2016. <http://arxiv.org/abs/1611.02351>.

Chris Whidden and Frederick A Matsen IV. Calculating the unrooted subtree prune-and-regraft distance. *arXiv*, 24 November 2015. <http://arxiv.org/abs/1511.07529>.

Duncan K Ralph and Frederick A Matsen IV. Likelihood-Based inference of B cell clonal families. *PLoS Comput. Biol.*, 12(10):e1005086, 17 October 2016.

Sara C Billey, Matjaž Konvalinka, and Frederick A Matsen IV. On the enumeration of tanglegrams and tangled chains. *J. Combin. Theory Ser. A*, 146:239–263, 2017.

Duncan K Ralph and Frederick A Matsen IV. Consistency of VDJ rearrangement and substitution parameters enables accurate B cell receptor sequence annotation. *PLOS Comput. Biol.*, 12(1):e1004409, January 2016.

Sarah Cobey, Patrick Wilson, and Frederick A Matsen IV. The evolution within us. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 370(1676), 5 September 2015.

Connor O McCoy, Trevor Bedford, Vladimir N Minin, Philip Bradley, Harlan Robins, and Frederick A Matsen IV. Quantifying evolutionary constraints on B-cell affinity maturation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 370(1676), 5 September 2015.

Chris Whidden and Frederick A Matsen IV. Quantifying MCMC exploration of phylogenetic tree space. *Systematic Biology*, 64(3):472–491, May 2015.

Alevtina Gall, Jutta Fero, Connor McCoy, Brian C Claywell, Carissa A Sanchez, Patricia L Blount, Xiaohong Li, Thomas L Vaughan, Frederick A Matsen, Brian J Reid, and Nina R Salama. Bacterial composition of the human upper gastrointestinal tract microbiome is dynamic and associated with genomic instability in a Barrett’s esophagus cohort. *PLoS One*, 10(6):e0129055, 15 June 2015.

Dara A Lehman, Jared M Baeten, Connor O McCoy, Julie F Weis, Dylan Peterson, Gerald Mbara, Deborah Donnell, Katherine K Thomas, Craig W Hendrix, Mark A Marzinke, Lisa Frenkel, Patrick Ndase, Nelly R Mugo, Connie Celum, Julie Overbaugh, Frederick A Matsen, and The Partners PrEP study team. Risk of drug resistance among persons acquiring HIV within a randomized clinical trial of single- or Dual-Agent preexposure prophylaxis. *J. Infect. Dis.*, 13 January 2015.

Frederick Albert Matsen IV

Frederick A Matsen IV. Phylogenetics and the human microbiome. *Syst. Biol.* (invited review), 64(1):e26–41, January 2015.

Noah G Hoffman and Frederick A Matsen. Computational tools for taxonomic assignment. In *Encyclopedia of Metagenomics*, pages 1–6. Springer New York, 3 April 2014.

Aaron E Darling, Guillaume Jospin, Eric Lowe, Frederick A Matsen IV, Holly M Bik, and Jonathan A Eisen. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*, 2:e243, 9 January 2014. Selected as one of the top 20 articles from 2014 in PeerJ (<http://peerj.com/collections/5-peerjpicks/>).

Frederick A Matsen IV, Christopher T Small, Khanh Soliven, Gregory A Engel, Mostafa M Feeroz, Xiaoxing Wang, Karen L Craig, M Kamrul Hasan, Michael Emerman, Maxine L Linial, and Lisa Jones-Engel. A novel Bayesian method for detection of APOBEC3-Mediated hypermutation and its application to zoonotic transmission of simian foamy viruses. *PLoS Comput. Biol.*, 10(2):e1003493, 27 February 2014.

Khanh Soliven, Xiaoxing Wang, Christopher T Small, Mostafa M Feeroz, Eun-Gyung Lee, Karen L Craig, Kamrul Hasan, Gregory A Engel, Lisa Jones-Engel, Frederick A Matsen, and Maxine L. Linial. Simian foamy virus infection of rhesus macaques in Bangladesh: Relationship of latent proviruses and transcriptionally active viruses. *Journal of virology*, 87(24):13628–13639, 2013.

L. Jones-Engel, G. Engel, C. Small, K. Soliven, M. Feeroz, X. Wang, M. Hasan, G. Oh, S. M. R. Alam, K. Craig, D. Jackson, F. Matsen IV, and M. L. Linial. Zoonotic simian foamy virus in Bangladesh reflects diverse patterns of transmission and co-infection. *Emerging Microbes and Infections*, 2(9):e58, 2013.

K. Ronen, C. O. McCoy, F. A. Matsen, D. F. Boyd, S. Emery, K. Odem-Davis, W. Jaoko, K. Mandaliya, R. S. McClelland, B. A. Richardson, and J. Overbaugh. HIV-1 superinfection occurs less frequently than initial infection in a cohort of high-risk Kenyan women. *PLOS Pathogens*, 9(8):e1003593, 2013.

L. Etienne, B.H. Hahn, P.M. Sharp, F.A. Matsen, and M. Emerman. Gene loss and adaptation to hominoids underlie the ancient origin of HIV-1. *Cell Host & Microbe*, 14(1):85–92, 2013.

C.O. McCoy and F.A. Matsen. Abundance-weighted phylogenetic diversity measures distinguish microbial community states and are robust to sampling depth. *PeerJ*, 9:e157, 2013.

E. H. Warren, F.A. Matsen, and J. Chou. High-throughput sequencing of B- and T-lymphocyte antigen receptors in hematology. *Blood*, 2013.

M.M. Feeroz, K. Soliven, C.T. Small, G.A. Engel, M.A. Pacheco-Delgado, J.L. Yee, X. Wang, M. Kamrul Hasan, G. Oh, K.L. Levine, S.M.R. Alam, K. L. Craig, D.L. Jackson, E. Lee, P.A. Barry, N.W. Lerche, A.A. Escalante, F.A. Matsen, M.L. Linial, and L. Jones-Engel. Population dynamics of rhesus macaques and associated foamy virus in Bangladesh. *Emerging Microbes and Infections*, 2013.

D.A. Nipperess and F.A. Matsen. The mean and variance of phylogenetic diversity under rarefaction. *Methods in Ecology and Evolution*, 4(6):566–572, 2013.

F.A. Matsen, A. Gallagher, and C. McCoy. Minimizing the average distance to a closest leaf in a phylogenetic tree. *Systematic Biology*, 62(6):824–836, November 2013.

S. J. Salipante, D. J. Sengupta, C. Rosenthal, G. Costa, J. Spangler, E. H. Sims, M. A. Jacobs, S. I. Miller, D. R. Hoogstraal, B. T. Cookson, C. O. McCoy, F. A. Matsen, J. Shendure, C. C. Lee, T. T. Harkins, and N. G. Hoffman. Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. *PLOS ONE*, 8(5):e65226, 2013.

F.A. Matsen and S.N. Evans. Edge principal components and squash clustering: using the special structure of phylogenetic placement data for sample comparison. *PLOS ONE*, 2012.

- F.A. Matsen, A. Gallagher, and C. McCoy. nestly—a framework for running software with nested parameter choices and aggregating results. *Bioinformatics*, 29(3):387–388, 2013.
- M.T. Levine, C. McCoy, D. Vermaak, Y.C.G. Lee, M.A. Hiatt, F.A. Matsen, and H.S. Malik. Phylogenomic analysis reveals dynamic evolutionary history of the *Drosophila* heterochromatin protein 1 (HP1) gene family. *PLOS Genetics*, 8(6):e1002729, 2012.
- S. Srinivasan, N.G. Hoffman, M.T. Morgan, F.A. Matsen, T.L. Fiedler, R.W. Hall, F.J. Ross, C.O. McCoy, R. Bumgarner, J.M. Marrazzo, and D.N. Fredricks. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLOS ONE*, 7(6):e37818, 2012.
- D.A. Lehman, D.C. Wamalwa, C.O. McCoy, F.A. Matsen, A. Langat, B.H. Chohan, S. Benki-Nugent, R. Custers-Allen, F.D. Bushman, G.C. John-Stewart, and J. Overbaugh. Low-frequency nevirapine resistance at multiple sites may predict treatment failure in infants on nevirapine-based treatment. *Journal of Acquired Immune Deficiency Syndromes*, 60(3):225, 2012.
- S. N. Evans and F. A. Matsen. The phylogenetic Kantorovich-Rubinstein metric for environmental sequence samples. *Journal of the Royal Statistical Society (B)*, 74(3):569–592, 2012.
- F.A. Matsen, N.G. Hoffman, A. Gallagher, and A. Stamatakis. A format for phylogenetic placements. *PLOS ONE*, 7(2):e31009, 2012.
- E.S. Lim, O.I. Fregoso, C.O. McCoy, F.A. Matsen, H.S. Malik, and M. Emerman. The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host and Microbe*, 11(2):194–204, 2012.
- F. A. Matsen and A. Gallagher. Reconciling taxonomy and phylogenetic inference: formalism and algorithms for describing discord and inferring taxonomic roots. *Algorithms for Molecular Biology*, 7:8, 2012.
- F.A. Matsen and S.N. Evans. Ubiquity of synonymity: almost all large binary trees are not uniquely identified by their spectra or their immanantal polynomials. *Algorithms for Molecular Biology*, 7:14, 2012.
- F.A. Matsen. constNJ: an algorithm to reconstruct sets of phylogenetic trees satisfying pairwise topological constraints. *Journal of Computational Biology*, 17(6):799–818, 2010.
- M.A. Cueto and F.A. Matsen. Polyhedral geometry of phylogenetic rogue taxa. *Bulletin of Mathematical Biology*, 73(6):1202–1226, 2011.
- F.A. Matsen, R. Kodner, and E.V. Armbrust. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1):538, 2010.
- F.A. Matsen. Fourier transform inequalities for phylogenetic trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 6(1):89–95, 2009.
- D. Ford, F.A. Matsen, and T. Stadler. A method for investigating relative timing information on phylogenetic trees. *Systematic Biology*, 58(2):167–183, 2009.
- F.A. Matsen and S.N. Evans. To what extent does genealogical ancestry imply genetic ancestry? *Theoretical Population Biology*, 74(2):182–190, 2008.
- F.A. Matsen, E. Mossel, and M. Steel. Mixed-up trees: the structure of phylogenetic mixtures. *Bulletin of Mathematical Biology*, 70(4):1115–1139, 2008.
- F. Matsen. Optimization over a class of tree shape statistics. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 4(3):506–512, 2007.
- F.A. Matsen and M. Steel. Phylogenetic mixtures on a single tree can mimic a tree of another topology. *Systematic Biology*, 56(5):767–775, 2007.
- M. Steel and F.A. Matsen. The Bayesian “star paradox” persists for long finite sequences. *Molecular Biology and Evolution*, 24(4):1075–1079, 2007.

Frederick Albert Matsen IV

F.A. Matsen and J. Wakeley. Convergence to the island-model coalescent process in populations with restricted migration. *Genetics*, 172(1):701, 2006.

F.A. Matsen. A geometric approach to tree shape statistics. *Systematic Biology*, 55(4):652–661, 2006.

F.A. Matsen and M.A. Nowak. Win–stay, lose–shift in language learning from peers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(52):18053, 2004.